

# Convergence rates and source conditions for Tikhonov regularization with sparsity constraints

Dirk A. Lorenz\*

July 18, 2008

## Abstract

This paper addresses the regularization by sparsity constraints by means of weighted  $\ell^p$  penalties for  $0 \leq p \leq 2$ . For  $1 \leq p \leq 2$  special attention is paid to convergence rates in norm and to source conditions. As main results it is proven that one gets a convergence rate of  $\sqrt{\delta}$  in the 2-norm for  $1 < p \leq 2$  and in the 1-norm for  $p = 1$  as soon as the unknown solution is sparse. The case  $p = 1$  needs a special technique where not only Bregman distances but also a so-called Bregman-Taylor distance has to be employed.

For  $p < 1$  only preliminary results are shown. These results indicate that, different from  $p \geq 1$ , the regularizing properties depend on the interplay of the operator and the basis of sparsity. A counterexample for  $p = 0$  shows that regularization need not to happen.

**AMS Subject classification:** Primary 47A52; Secondary 65J20, 65F22.

## 1 Introduction

In this paper we discuss the regularizing properties of so-called sparsity constraints. We consider linear inverse problems with a bounded operator  $A : X \rightarrow Y$  between two Hilbert spaces. Our setting is classical [12]: We assume that we are given noisy data  $g^\delta \in Y$  such that there exists  $g^+ = Af^+$  with  $\|g^+ - g^\delta\|_Y \leq \delta$ . Our aim is to reconstruct  $f^+$  from the noisy data  $g^\delta$ . It is well known that this problem is ill-posed if and only if the range of  $A$  is non-closed [12].

Recently regularization with sparsity constraints has become popular due to the influential paper [9]. In this setting one assumes, that the unknown solution has a sparse representation in a certain orthonormal basis or frame  $(\psi_k)$  of  $X$ , i.e. the unknown solution  $f^+$  can be expressed as  $f^+ = \sum u_k \psi_k$  where the sum consists of a few (and especially finitely many) terms only. This knowledge is used to set up a so-called sparsity constraint for Tikhonov regularization, i.e. the regularized solution is given as a minimizer of

$$\|Af - g^\delta\|_Y^2 + \alpha \sum_k w_k \phi(|\langle f | \psi_k \rangle|)$$

---

\*Dirk Lorenz, Zentrum für Technomathematik, Fachbereich 3, Universität Bremen, PO Box 330440, 28334 Bremen, Germany, [dlorenz@math.uni-bremen.de](mailto:dlorenz@math.uni-bremen.de).

with a suitably chosen function  $\phi$ . The parameter  $\alpha > 0$  is a regularization parameter and the weighting sequence  $w_k > 0$  allows to regularize each coefficient individually. For the weighting sequence we assume that it is bounded away from zero:  $w_k \geq w_0 > 0$ . Several choices of  $\phi$  are possible. In [9] it is argued that the choice  $\phi(s) = s^p$  for  $1 \leq p < 2$  promotes sparsity of the minimizer. A heuristic explanation is that this functions give a higher weight to small coefficients and lower weight to large coefficients. Of course the cases  $p < 1$  or even  $p = 0$  will produce sparse minimizers but in this case the convexity of the functional is lost and minimizers need not to exist (see [16] for a discussion of the case  $A = I$ ).

For notational convenience we introduce the synthesis operator  $B : \ell^2 \rightarrow X$  defined by  $Bu = \sum_k u_k \psi_k$ . We define  $K = AB$  and rewrite the Tikhonov functional as

$$\Psi(u) = \|Ku - g^\delta\|_Y^2 + \alpha \sum w_k |u_k|^p. \quad (1)$$

The calculation of a minimizer of the above functional is not a straightforward task. Convergent algorithms in the infinite dimensional setting for  $1 \leq p \leq 2$  were proposed and analyzed in [4–6, 9, 10, 14]. Generalizations to joint sparsity [13], nonlinear operators [2, 17, 18] and the case  $p = 0$  [1] have been proposed.

In this paper we are going to discuss the regularizing properties of sparsity constraints. First results on this topic can be found in [9] where convergence of the minimizers in  $X$  (resp  $\ell^2$ ) for vanishing noise and the parameter choice  $\alpha(\delta)$  such that  $\alpha \rightarrow 0$  and  $\delta^2/\alpha \rightarrow 0$  has been shown. Moreover, it is shown that, in the special case of wavelet bases with a special class of weights which lead to Besov spaces, convergence rates can be achieved. The paper [18] also deals with convergence of the minimizers and the proofs there show that convergence in the stronger  $\ell^1$  norm holds. Sparsity constraints can also be discussed in the framework of regularization in Banach spaces like, e.g., in [7, 8, 15, 19, 20]. In these papers convergence rates for general convex regularization are given in terms of Bregman distances. In this paper we focus on convergence rates for sparsity constraints in norm, i.e. in the norm in  $X$  resp.  $\ell^2$  or the  $\ell^1$ -norm.

The paper is organized as follows. Section 2 presents auxiliary results and in Section 3 results on convergence rates for Tikhonov regularization with (1) for  $1 < p \leq 2$  are presented, especially we illustrate the role of the source condition. Section 4 treats the case  $p = 1$  which is considerably different and a different technique has to be used. The Section 5 collects preliminary results on the regularization with  $p < 1$ . Here, no convergence rates can be given so far, and are not to be expected in general. In the last section we draw conclusions.

**Notation.** We denote with  $\ell_w^p$  the weighted  $\ell^p$  space, i.e. the sequences  $u$  such that  $\sum w_k |u_k|^p$  converges. We consider the spaces  $\ell_w^p$  for  $0 < p < \infty$  which are normed spaces (quasi-normed for  $p < 1$ ) when equipped with the (quasi-)norm  $\|u\|_{p,w} = (\sum w_k |u_k|^p)^{1/p}$ . By  $\ell^0$  we denote the set  $\{u : \mathbf{N} \rightarrow \mathbf{R} : u_k \neq 0 \text{ for finitely many } k\}$  of finitely supported or sparse sequences and with  $\ell_w^0$  the set  $\{u : \mathbf{N} \rightarrow \mathbf{R} : \sum w_k \operatorname{sgn}(|u_k|) < \infty\}$ . For simplicity we write  $\|u\| = \|u\|_2$  and the inner product of  $u, v \in \ell^2$  is denoted by  $\langle u | v \rangle$ . Moreover, we will frequently use component-wise application of operators to sequences,

e.g.  $(|u|^p)_k = |u_k|^p$  or  $(wu)_k = w_k u_k$ . With

$$\text{Sgn}(x) = \begin{cases} \{1\} & \text{for } x > 0 \\ [-1, 1] & \text{for } x = 0 \\ \{-1\} & \text{for } x < 0. \end{cases}$$

we denote the multivalued sign while  $\text{sgn}$  stands for the usual sign with  $\text{sgn}(0) = 0$ . For an operator  $A : X \rightarrow Y$  between two Hilbert spaces the Hilbert space adjoint is denoted by  $A^* : Y \rightarrow X$ .

## 2 Preliminary results

In this section we collect preliminary results which are needed in the following.

As a first result we report that the cases  $1 \leq p < 2$  indeed promote sparsity and that  $p = 1$  lead to finitely supported minimizers.

**Lemma 2.1.** *Let  $1 \leq p \leq 2$ . A minimizer  $u^*$  of  $\Psi$  from (1) fulfills*

$$u^* \in \ell_{w^2}^{2(p-1)}.$$

*Proof.* Every minimizer  $u$  of  $\Phi$  fulfills

$$-2K^*(Ku - g^\delta) \in \alpha w p \text{Sgn}(u)|u|^{p-1}. \quad (2)$$

For  $p > 1$  the inclusion becomes an equation and since the left hand side is an  $\ell^2$  sequence, the right hand side is also in  $\ell^2$ . It follows that

$$\sum w_k^2 |u_k|^{2(p-1)} < \infty.$$

For  $p = 1$  assume that  $u \notin \ell_{w^2}^0$  i.e. the sum  $\sum w_k^2 \text{sgn}(|u_k|)$  diverges. Hence, every other choice of a sign in (2) also leads to a diverging sum and it follows that the left hand side in (2) can not be an  $\ell^2$  sequence, which is a contradiction.  $\square$

The next statement is on convergence of minimizers of (1) for  $\delta \rightarrow 0$ .

**Theorem 2.2** ([9]). *Assume that either  $p > 1$  or  $K$  is injective,  $w_k \geq w_0 > 0$ , and let  $u^{\alpha, \delta}$  be a minimizer of  $\Psi$  from (1). If the parameter choice  $\alpha(\delta)$  fulfills*

$$\lim_{\delta \rightarrow 0} \alpha(\delta) = 0, \quad \lim_{\delta \rightarrow 0} \frac{\delta^2}{\alpha(\delta)} = 0$$

*then it holds*

$$\lim_{\delta \rightarrow 0} \|u^{\alpha, \delta} - u^+\| = 0.$$

This says that that the method is indeed a regularization. To get a statement on the rate of convergence the true solution  $u^+$  has to fulfill some source condition. This will be topic of sections 3 and 4.

Next we state a basic inequality which we will need in the following.

**Lemma 2.3** ([4]). *Let  $1 < p \leq 2$ . For  $C > 0$  and  $L > 0$  it holds for every  $s, t \in \mathbf{R}$  with  $|s| \leq C$  and  $|t - s| \leq L$*

$$|t|^p - |s|^p \geq p \text{sgn}(s)|s|^{p-1}(t - s) + \kappa |t - s|^2$$

*with  $\kappa = \frac{p(p-1)}{2(C+L)^{2-p}}$ .*

### 3 Regularization with $1 < p \leq 2$

In this section we analyze the “easiest” case  $1 < p \leq 2$ . The main result goes as follows.

**Theorem 3.1.** *Let  $1 < p \leq 2$ ,  $w_k \geq w_0 > 0$  and let  $u^{\alpha,\delta}$  be a minimizer of  $\Psi$  given in (1). Furthermore let  $u^+$  fulfill the source condition*

$$\exists \theta \in Y : w \operatorname{sgn}(u^+) |u^+|^{p-1} = K^* \theta. \quad (3)$$

Then for the choice  $\alpha \sim \delta$  it holds

$$\|Ku^{\alpha,\delta} - g^\delta\|_Y = \mathcal{O}(\delta) \quad \text{for } \delta \rightarrow 0 \quad (4)$$

$$\sum w_k |u_k^{\alpha,\delta} - u_k^+|^2 = \mathcal{O}(\delta) \quad \text{for } \delta \rightarrow 0. \quad (5)$$

*Proof.* Due to the minimizing property we have

$$\|Ku^{\alpha,\delta} - g^\delta\|_Y^2 + \alpha \sum w_k |u_k^{\alpha,\delta}|^p \leq \delta^2 + \alpha \sum w_k |u_k^+|^p$$

which gives

$$\|Ku^{\alpha,\delta} - g^\delta\|_Y^2 + \alpha \sum w_k (|u_k^{\alpha,\delta}|^p - |u_k^+|^p) \leq \delta^2.$$

Since  $|u_k^+|$  and  $|u_k^{\alpha,\delta} - u_k^+|$  can be bounded uniformly in  $k$  (the second due to Theorem 2.2) we can apply Lemma 2.3 which yields

$$\|Ku^{\alpha,\delta} - g^\delta\|_Y^2 + \alpha \kappa \sum w_k |u_k^{\alpha,\delta} - u_k^+|^2 + p\alpha \sum w_k \operatorname{sgn}(u_k^+) |u_k^+|^{p-1} (u_k^{\alpha,\delta} - u_k^+) \leq \delta^2.$$

Rearranging gives

$$\|Ku^{\alpha,\delta} - g^\delta\|_Y^2 + \alpha \kappa \sum w_k |u_k^{\alpha,\delta} - u_k^+|^2 \leq \delta^2 + \alpha \langle pw \operatorname{sgn}(u^+) |u^+|^{p-1} |u^+ - u^{\alpha,\delta} \rangle.$$

Applying the source condition (3) and the Cauchy-Schwarz inequality leads to

$$\|Ku^{\alpha,\delta} - g^\delta\|_Y^2 + \alpha \kappa \sum w_k |u_k^{\alpha,\delta} - u_k^+|^2 \leq \delta^2 + \alpha p \|\theta\|_Y \|K(u^+ - u^{\alpha,\delta})\|_Y.$$

Adding and subtracting  $g^\delta$  in the last norm and denoting  $\rho = \|\theta\|_Y p/2$  leads to

$$\|Ku^{\alpha,\delta} - g^\delta\|_Y^2 + \alpha \kappa \sum w_k |u_k^{\alpha,\delta} - u_k^+|^2 \leq \delta^2 + 2\alpha\rho\delta + 2\alpha\rho \|Ku^{\alpha,\delta} - g^\delta\|_Y.$$

Rearranging and completing the squares gives

$$(\|Ku^{\alpha,\delta} - g^\delta\|_Y - \alpha\rho)^2 + \alpha \kappa \sum w_k |u_k^{\alpha,\delta} - u_k^+|^2 \leq (\delta + \alpha\rho)^2.$$

This finally implies

$$\|Ku^{\alpha,\delta} - g^\delta\|_Y \leq \delta + 2\alpha\rho \quad (6)$$

and

$$\sum w_k |u_k^{\alpha,\delta} - u_k^+|^2 \leq \frac{(\delta + \alpha\rho)^2}{\alpha \kappa}. \quad (7)$$

The assertion follows with  $\alpha \sim \delta$ .  $\square$

Since  $w_k \geq w_0$  we can deduce the following corollary immediately.

**Corollary 3.2.** *Under the assumptions of Theorem 3.1 it holds*

$$\|u^{\alpha,\delta} - u^+\| = \mathcal{O}(\sqrt{\delta}).$$

We state a few remarks to illustrate Theorem 3.1.

**Remark 3.3** (Constants in the  $\mathcal{O}$ -notation). *From (6) one deduces that*

$$\|Ku^{\alpha,\delta} - g^\delta\|_Y \leq (1 + 2\rho)\delta$$

and hence, the constant in the  $\mathcal{O}$  notation only depends on  $\rho$ . From the estimate (7) we have

$$\sum w_k |u_k^{\alpha,\delta} - u_k^+|^2 \leq \frac{(1 + \rho)^2}{\kappa} \delta.$$

In this case, the constant depends also on  $\kappa$  from Lemma 2.3 for which it holds

$$\frac{1}{\kappa} = \frac{2(C + L)^{2-p}}{p(p-1)}$$

where  $C$  is an upper bound on  $|u_k^+|$  and  $L$  is an upper bound for  $|u_k^{\alpha,\delta} - u_k^+|$ . The value  $L$  tends to zero for  $\delta \rightarrow 0$  and  $C$  depends on  $u^+$  only and hence,  $C$  and  $L$  are uniformly bounded for  $\delta \rightarrow 0$ . Finally, we see that the constant  $1/\kappa$  mainly depends on  $p$  and is large for small  $p$  and namely it tends to infinity for  $p \rightarrow 1$ . To summarize, we may say that the regularization with a weighted  $\ell^p$ -norm leads to a convergence rate of order  $\sqrt{\delta}$  in the 2-norm but the associated constant gets arbitrarily large for  $p$  close to one. Hence, one may not assume a similar theorem to hold for the limiting case  $p = 1$ . Fortunately, Theorem 4.3 below shows that this pessimism unfounded.

**Remark 3.4** (The results of Burger and Osher [7]). *In the case of a general convex and lower-semicontinuous penalty functional  $J$ , Burger and Osher proved that the source condition*

$$\exists \theta : K^* \theta \in \partial J(u^+)$$

leads to a convergence rate

$$D_\xi(u^{\alpha,\delta}, u^+) = \mathcal{O}(\delta)$$

for  $u^{\alpha,\delta}$  minimizers of

$$\|Ku - g^\delta\|_Y + \alpha J(u).$$

Here  $\partial J$  denotes the subgradient of  $J$ ,  $\xi \in \partial J(u^+)$  and

$$D_\xi(u^{\alpha,\delta}, u^+) = J(u^{\alpha,\delta}) - J(u^+) - \langle \xi | u^{\alpha,\delta} - u^+ \rangle$$

is the Bregman distance, see also [15]. One can also deduce Theorem 3.1 from this result by noting that this source condition is precisely the one in Theorem 3.1 and that for  $1 < p \leq 2$  the Bregman distance of  $J(u) = \sum w_k |u_k|^p$  can be bounded from below:

$$\sum w_k |u^{\alpha,\delta} - u^+|^2 \leq D_\xi(u^{\alpha,\delta}, u^+)$$

for  $\|u^{\alpha,\delta} - u^+\| < M$  (which follows from Lemma 2.3 or the inequalities of Xu and Roach, see [21–23]).

**Remark 3.5** (Source conditions in terms of  $\ell^p$ -spaces). *In the classical (quadratic) theory the source condition can usually be interpreted as some kind of smoothness condition. When working in sequence space, we see that the source condition (3) says something about the decay of the solution  $u^+$ . We assume that the operator under consideration has the property  $\text{range } K^* = \ell_v^q$  where we assume that the space  $\ell_v^q$  is contained in  $\ell^2$ . Hence, the dual space  $(\ell_v^q)' = \ell_{v'}^{q'}$  with dual exponent  $q' = q/(q-1)$  and dual weight  $v^{-1/(q-1)}$  is larger than  $\ell^2$ . One may say that the operator  $K : \ell_{v'}^{q'} \rightarrow Y$  has a “smoothing” (or better “damping”) property. Now, the source condition (3) reads as  $w \text{sgn}(u^+) |u^+|^{p-1} \in \text{range } K^* = \ell_v^q$  and hence*

$$\sum v_k w_k^q |u_k^+|^{q(p-1)} < \infty \quad \text{or equivalently} \quad u^+ \in \ell_{vw^q}^{q(p-1)}.$$

## 4 Regularization with $p = 1$

We now turn to the case  $p = 1$ . In this case previous results give convergence rates in the Bregman distance only [7, 15, 19, 20]. Moreover, Remark 3.4 does not apply, since the function  $J(u) = \sum_k w_k |u_k|$  is not strictly convex and hence, the Bregman distance with respect to the functional  $J(u) = \sum w_k |u_k|$  can not be estimated by the  $\ell^2$ -norm in general. It holds  $\partial J(u) = (w_k \text{Sgn}(u_k))_k$ . One sees that the Bregman distance fulfills

$$D_\xi(u, u^+) \leq 2 \sum_{\substack{(u_k > 0 \wedge u_k \leq 0) \\ \vee (u_k < 0 \wedge u_k \geq 0)}} |u_k|.$$

Consequently, the Bregman distance is zero as soon as the signs of  $u$  and  $u^+$  coincide and a convergence rate regarding the Bregman distance does not give satisfactory information, see also [8].

To prove a convergence rate like in Theorem 3.1 we need the following lemma which can be found in similar form in [5]. As an important ingredient we need the so called FBI property, also from [5].

**Definition 4.1.** *An operator  $K : \ell^2 \rightarrow Y$  mapping into a Hilbert space has the finite basis injectivity (FBI) property, if for all finite subsets  $I \subset \mathbf{N}$  the operator  $K|_I$  is injective, i.e. for all  $u, v \in \ell^2$  with  $Ku = Kv$  and  $u_k = v_k = 0$  for all  $k \notin I$  it follows  $u = v$ .*

The lemma gives an estimate which compares the Bregman distance with the  $\ell^1$ -norm.

**Lemma 4.2.** *Let  $u^+$  have finite support,  $w_k \geq w_0 > 0$ , let  $K$  fulfill the FBI property, and define*

$$T(u) = \|K(u - u^+)\|_Y^2 \tag{8}$$

$$R(u) = \sum w_k |u_k| - \sum w_k |u_k^+| - \sum w_k \text{sgn}(u_k^+) (u_k - u_k^+). \tag{9}$$

*Then there exists  $\lambda > 0$  such that*

$$R(u) + T(u) \geq \lambda \|u - u^+\|_1^2$$

*whenever  $\|u - u^+\|_1 \leq M$ .*

*Proof.* We define  $I = \{k \mid \text{sgn}(u_k^+) = \pm 1\}$  which is a finite set. We estimate

$$\begin{aligned} R(u) &= \sum_k w_k |u_k| - w_k |u_k^+| - w_k \text{sgn}(u_k^+)(u_k - u_k^+) \\ &= \sum_k w_k |u_k| - w_k \text{sgn}(u_k^+) u_k \\ &\geq \sum_{k \notin I} w_k |u_k| - w_k \text{sgn}(u_k^+) u_k = \sum_{k \notin I} w_k |u_k|. \end{aligned}$$

Denoting with  $I^c$  the complement of  $I$  and with  $P_{I^c}$  the projection onto the subspace where all coefficients in  $I$  are zero we get (using  $u_k^+ = 0$  for  $k \in I^c$ )

$$R(u) \geq w_0 \|P_{I^c}(u - u^+)\|_1.$$

Since  $\|P_{I^c}(u - u^+)\| \leq M$  we can estimate

$$R(u) \geq \frac{w_0}{M} \|P_{I^c}(u - u^+)\|_1^2. \quad (10)$$

To establish an estimate for the remaining part  $P_I u$  we start with  $u = P_I u + P_{I^c} u$  and use the inequalities of Cauchy-Schwarz (in the form  $-\langle u|v \rangle \leq \|u\| \|v\|$ ) and Young ( $ab \leq \frac{a^2}{4} + b^2$  for  $a, b > 0$ ) to get

$$\begin{aligned} \|Ku\|_Y^2 &= \|KP_I u\|_Y^2 + 2 \langle KP_I u | KP_{I^c} u \rangle + \|KP_{I^c} u\|_Y^2 \\ &\geq \frac{\|KP_I u\|_Y^2}{2} - \|KP_{I^c} u\|_Y^2 \\ &\geq \frac{\|KP_I u\|_Y^2}{2} - \|K\|^2 \|P_{I^c} u\|^2. \end{aligned} \quad (11)$$

Since  $I$  is finite and  $K$  obeys the FBI property there is a constant  $c > 0$  such that

$$c \|P_I u\|^2 \leq \|KP_I u\|_Y^2.$$

Moreover, again since  $I$  is finite, we can estimate the 2-norm from below by the 1-norm which leads to

$$\tilde{c} \|P_I u\|_1^2 \leq \|KP_I u\|_Y^2.$$

Combining this with (11) gives

$$\|P_I u\|_1^2 \leq \frac{2}{\tilde{c}} (\|Ku\|_Y^2 + \|K\|^2 \|P_{I^c} u\|^2)$$

Applying this estimate to  $u - u^+$  instead of  $u$  and adding the inequality (10) leads to

$$\|u - u^+\|_1^2 \leq \frac{2}{\tilde{c}} (T(u) + \|K\|^2 \|P_{I^c}(u - u^+)\|^2) + \frac{M}{w_0} R(u).$$

By estimating the 1-norm from below by the 2-norm in (10) we get  $\frac{M}{w_0} R(u) \geq \|P_{I^c}(u - u^+)\|_2^2$  and hence,

$$\|u - u^+\|_1^2 \leq \frac{2}{\tilde{c}} T(u) + \frac{M}{w_0} \left( \frac{2\|K\|^2}{\tilde{c}} + 1 \right) R(u)$$

which proves the claim.  $\square$

While the term  $R$  from (9) is a Bregman distance, the term  $T$  from (8) can be seen as Taylor distance: We define the functional  $F(u) = \|Ku - g^\delta\|_Y^2$  and observe that the term  $T$  can be rewritten as

$$T(u) = F(u) - F(u^+) - \langle F'(u^+) | u - u^+ \rangle.$$

Consequently,  $T$  is the remainder of the Taylor expansion of the fidelity term  $F$ . Therefore, Lemma 4.2 can be seen as an estimate on the Bregman-Taylor-distance  $R + T$ .

Lemma 4.2 enables us to prove the main result of this paper:

**Theorem 4.3.** *Let  $u^+$  have finite support,  $w_k \geq w_0 > 0$ ,  $K$  obey the FBI property, and let furthermore  $u^+$  fulfill the source condition*

$$\exists \theta \in Y : w \operatorname{sgn}(u^+) = K^* \theta. \quad (12)$$

Then for every

$$u^{\alpha, \delta} \in \operatorname{argmin} \|Ku - g^\delta\|_Y^2 + \alpha \sum w_k |u_k|$$

it holds

$$\|u^{\alpha, \delta} - u^+\|_1 = \mathcal{O}(\sqrt{\delta}).$$

*Proof.* Due to the minimizing property we have

$$\begin{aligned} 0 &\leq \|Ku^+ - g^\delta\|_Y^2 + \alpha \sum_k w_k |u_k^+| - \|Ku^{\alpha, \delta} - g^\delta\|_Y^2 - \alpha \sum_k w_k |u_k^{\alpha, \delta}| \\ &= \|Ku^+ - g^\delta\|_Y^2 - \|Ku^{\alpha, \delta} - g^\delta\|_Y^2 \\ &\quad + \alpha \left( \sum_k w_k |u_k^+| - \sum_k w_k |u_k^{\alpha, \delta}| + \sum_k w_k \operatorname{sgn}(u_k^+) (u_k^{\alpha, \delta} - u_k^+) \right) \\ &\quad - \alpha \sum_k w_k \operatorname{sgn}(u_k^+) (u_k^{\alpha, \delta} - u_k^+). \end{aligned}$$

Rearranging gives

$$\alpha R(u^{\alpha, \delta}) \leq \delta^2 - \|Ku^{\alpha, \delta} - g^\delta\|_Y^2 - \alpha \sum_k w_k \operatorname{sgn}(u_k^+) (u_k^{\alpha, \delta} - u_k^+).$$

Since the convergence  $u^{\alpha, \delta} \rightarrow u^+$  is known from Theorem 2.2 we can use Lemma 4.2 to obtain

$$\alpha \lambda \|u^{\alpha, \delta} - u^+\|_1^2 - \alpha \|K(u^{\alpha, \delta} - u^+)\|_Y^2 \leq \delta^2 - \|Ku^{\alpha, \delta} - g^\delta\|_Y^2 - \alpha \sum_k w_k \operatorname{sgn}(u_k^+) (u_k^{\alpha, \delta} - u_k^+).$$

With the source condition (12), the notation  $\rho = \|\theta\|_Y/2$ , and the Cauchy-Schwarz inequality this gives

$$\alpha \lambda \|u^{\alpha, \delta} - u^+\|_1^2 - \alpha \|K(u^{\alpha, \delta} - u^+)\|_Y^2 \leq \delta^2 - \|Ku^{\alpha, \delta} - g^\delta\|_Y^2 + \alpha 2\rho \|K(u^{\alpha, \delta} - u^+)\|.$$

Adding and subtracting  $g^\delta$  in the last norm and rearranging leads to

$$\alpha \lambda \|u^{\alpha, \delta} - u^+\|_1^2 - \alpha \|K(u^{\alpha, \delta} - u^+)\|_Y^2 + \|Ku^{\alpha, \delta} - g^\delta\|_Y^2 - 2\alpha \rho \|Ku^{\alpha, \delta} - g^\delta\|_Y \leq \delta^2 + 2\alpha \rho \delta.$$



Using

$$\|K(u^{\alpha,\delta} - u^+)\|_Y^2 \leq \|Ku^{\alpha,\delta} - g^\delta\|_Y^2 + 2\delta \|Ku^{\alpha,\delta} - g^\delta\|_Y + \delta^2$$

leads to

$$\alpha\lambda \|u^{\alpha,\delta} - u^+\|_1^2 + (1-\alpha) \|K(u^{\alpha,\delta} - g^\delta)\|_Y^2 - 2\alpha(\rho+\delta) \|Ku^{\alpha,\delta} - g^\delta\|_Y \leq (1+\alpha)\delta^2 + 2\alpha\rho\delta.$$

Dividing by  $(1-\alpha)$  and completing the square on the left hand side gives

$$\frac{\alpha}{1-\alpha} \lambda \|u^{\alpha,\delta} - u^+\|_1^2 + \left( \|Ku^{\alpha,\delta} - g^\delta\|_Y - \frac{\alpha}{1-\alpha}(\rho+\delta) \right)^2 \leq \frac{1+\alpha}{1-\alpha} \delta^2 + \frac{2\alpha\rho\delta}{1-\alpha} + \left( \frac{\alpha}{1-\alpha} \right)^2 (\rho+\delta)^2.$$

Finally, this gives

$$\begin{aligned} \|u^{\alpha,\delta} - u^+\|_1^2 &\leq \frac{1}{\lambda} \left( \frac{1+\alpha}{\alpha} \delta^2 + 2\rho\delta + \frac{\alpha}{1-\alpha} (\rho+\delta)^2 \right) \\ &= \frac{1}{\lambda\alpha(1-\alpha)} (\delta + \alpha\rho)^2. \end{aligned} \quad (13)$$

The choice  $\alpha = \delta$  proves

$$\|u^{\alpha,\delta} - u^+\|_1^2 = \mathcal{O}(\delta) \quad \text{for } \delta \rightarrow 0.$$

□

For  $p = 1$  the source condition says that  $u^+$  must only have a finite number of non-zero entries. This is the natural limit for  $p \rightarrow 1$  as can be seen from Remark 3.5.

Theorem 4.3 is remarkable since, as mentioned in Remark 3.3, the constant in the  $\mathcal{O}$ -notation in Theorem 3.1 blows up to infinity for  $p \rightarrow 1$ . Equation (13) shows that the constant in the  $\mathcal{O}$ -notation depends on the constant  $\lambda$  from Lemma 4.2 and on  $\rho = \|\theta\|_Y/2$  only. Basically the constant  $1/\kappa$  in Remark 3.3 has been replaced by  $1/\lambda$  from Lemma 4.2.

**Remark 4.4** (The result of Hofmann et al. [15]). *Hofmann et al. considered in [15] general convex regularization of operator equations in Banach spaces of the form*

$$\|F(u) - g^\delta\|_Y^p + \alpha J(u).$$

*They showed a convergence rate of  $\mathcal{O}(\delta)$  in the Bregman distance for non-smooth operators  $F$  under the source condition that there exists  $\beta_1 \in [0, 1]$ ,  $\beta_2 \geq 0$  and  $\xi \in \partial J(u^+)$  such that*

$$-\langle \xi | u - u^+ \rangle \leq \beta_1 D_\xi(u, u^+) + \beta_2 \|F(u) - F(u^+)\|$$

*(note that the negative sign on the left hand side is a typo in the original paper). This source condition is difficult to check in concrete situations. Applied to the situation of Theorem 4.3 it reads as: There exists  $\xi \in w\text{Sgn}(u^+)$  such that*

$$-\langle \xi | u - u^+ \rangle \leq \beta_1 D_\xi(u, u^+) + \beta_2 \|K(u - u^+)\|_Y.$$

*This condition is for example fulfilled if the sequence  $w_k$  is bounded and*

$$\|u - u^+\|_1 \leq \frac{1}{\max w_k} (\beta_1 R(u) + \beta_2 \|K(u - u^+)\|_Y)$$

*which resembles the Bregman-Taylor estimate from Lemma 4.2. However, Theorem 4.3 gives a convergence rate of  $\mathcal{O}(\sqrt{\delta})$  in the  $\ell^1$ -norm and the Bregman-Taylor estimate is only needed to pass from the Bregman distance to the  $\ell^1$ -norm. Additionally, Theorem 4.3 needs the source condition (12).*

## 5 Regularization with $p < 1$ ?

The functional (1) is not convex if  $p < 1$ . Hence, there is no guarantee for uniqueness or existence of a minimizer. In this section we show two extreme examples: One in which there exist minimizers which can be computed explicitly and regularization can be proven and the other where no minimizer exists at all.

### 5.1 Regularization is possible

In this example we use an orthonormal basis which is perfectly adapted to the operator: the singular basis. The singular value decomposition  $(\sigma_k, \psi_k, \phi_k)$  of the operator  $A$  consists of the singular values  $\sigma_k$  and two orthonormal bases  $\psi_k$  and  $\phi_k$  of  $X$  resp.  $Y$ . The operator  $A$  can now be expressed as

$$Af = \sum_k \sigma_k \langle f | \psi_k \rangle \phi_k.$$

Now we seek for a solution of  $Af = g$  which is sparse in the basis  $\psi_k$ , i.e. we have  $u_k = \langle f | \psi_k \rangle$  in (1). Hence, the operator  $K = AB$  has the form

$$Ku = ABu = \sum_k u_k \sigma_k \phi_k \quad (14)$$

To express the minimizer of (1) we need the following function:

$$H_\alpha^p(x) = \operatorname{argmin}_y (y - x)^2 + \alpha |y|^p. \quad (15)$$

Note that this function can be multivalued in general. The next lemma from [16] gives an implicit representation of the function  $H_\alpha^p$ .

**Lemma 5.1.** *Let*

$$G_\alpha^p(y) = y + \frac{\alpha p}{2} \operatorname{sgn}(y) |y|^{p-1}. \quad (16)$$

*The mapping  $H_\alpha^p$  is given by the following formulae:*

1. *Let  $1 < p \leq 2$ . Then  $(G_\alpha^p)^{-1}$  exists and is single valued and it holds*

$$H_\alpha^p(x) = (G_\alpha^p)^{-1}(x).$$

2. *Let  $p = 1$ . Then*

$$H_\alpha^1(x) = \max(|x| - \alpha/2, 0) \operatorname{sgn}(x).$$

3. *Let  $0 < p < 1$ . Then*

$$H_\alpha^p(x) = \begin{cases} 0 & , \text{ for } |x| \leq \alpha_{\text{eff}} \\ \text{the value of largest absolute value of the inverse} & , \text{ for } |x| \geq \alpha_{\text{eff}} \\ \text{mapping of } G_\alpha^p & \end{cases} \quad (17)$$

$$\text{where } \alpha_{\text{eff}} = \frac{2-p}{2-2p} \left( \alpha(1-p) \right)^{\frac{1}{2-p}}.$$

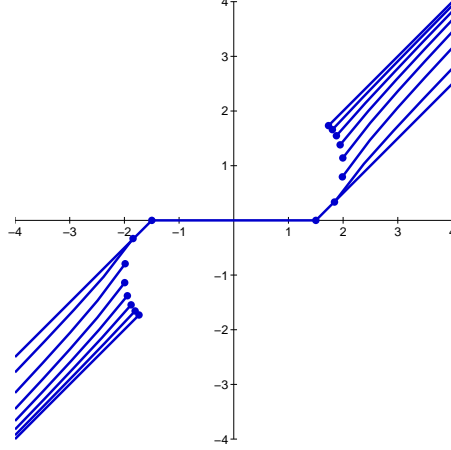


Figure 1: The thresholding functions  $H_\alpha^p$  for  $p = 0, 0.15, 0.3, 0.45, 0.6, 0.75, 0.9, 1$  and  $\alpha = 3$ .

4. Let  $p = 0$ . Then

$$H_\alpha^0(x) = \begin{cases} 0 & , \text{ for } |x| \leq \alpha_{\text{eff}} \\ x & , \text{ for } |x| \geq \alpha_{\text{eff}} \end{cases} \quad (18)$$

where  $\alpha_{\text{eff}} = \sqrt{\alpha}$ .

The description in 3. may be a little unfamiliar. For  $p < 1$  the function  $G_\alpha^p$  is multivalued in 0 with  $G_\alpha^p(0) = \mathbf{R}$ . Its inverse is again multivalued (in fact it has at most three values) and the function  $H_\alpha^p$  chooses either 0 or the value of largest absolute value, see Figure 1 and [16] for more details. Note moreover that for  $p < 1$  the function  $H_\alpha^p$  is multivalued itself, namely it has two values for  $|x| = \alpha_{\text{eff}}$ . For convenience we always choose the value 0 at these points in the following.

The next lemma is an easy consequence of the above lemma and the fact that the operator  $K$  is diagonal with respect to the basis  $(\phi_k)$  of  $Y$ .

**Lemma 5.2.** *Let  $(\phi_k)$  be an orthonormal basis of  $Y$  and let the operator  $K : \ell^2 \rightarrow Y$  be given by (14). Then, a minimizer of (1) is given by*

$$u_k^{\alpha, \delta} = \begin{cases} \frac{1}{\sigma_k} H_{\alpha/\sigma_k^p}^p(\langle g^\delta | \phi_k \rangle) & , \text{ for } \sigma_k > 0 \\ 0 & , \text{ for } \sigma_k = 0. \end{cases} \quad (19)$$

**Definition 5.3.** *For  $0 \leq p \leq 2$  we define the operator  $R_\alpha^p : Y \rightarrow \ell^2$  by*

$$(R_\alpha^p(g))_k = \begin{cases} \frac{1}{\sigma_k} H_{\alpha/\sigma_k^p}^p(\langle g^\delta | \phi_k \rangle) & , \text{ for } \sigma_k > 0 \\ 0 & , \text{ for } \sigma_k = 0. \end{cases}$$

Note that  $R_\alpha^p$  is non-linear and discontinuous.

**Theorem 5.4.** *Let  $0 \leq p < 1$ . The operator  $R_\alpha^p$  is*

1. defined for every  $g \in Y$ .

2. a regularization, i.e. for  $g \in \text{dom}(K^+)$  it holds

$$\lim_{\alpha \rightarrow 0} \|R_\alpha^p g - K^+ g\| = 0.$$

*Proof.* We abbreviate  $g_k = \langle g | \phi_k \rangle$ . The pseudo-inverse is given by

$$(K^+ g)_k = \begin{cases} g_k / \sigma_k & , \text{ for } \sigma_k > 0 \\ 0 & , \text{ for } \sigma_k = 0 \end{cases}$$

and by the Picard condition this is an  $\ell^2$  sequence. For an  $M \in \mathbf{N}$  we write

$$\begin{aligned} \|R_\alpha^p(g) - K^+ g\|^2 &= \sum_{\sigma_k > 0} \frac{|H_{\alpha/\sigma_k^p}^p(g_k) - g_k|^2}{\sigma_k^2} \\ &= \sum_{\sigma_k > 0, k \leq M} \frac{|H_{\alpha/\sigma_k^p}^p(g_k) - g_k|^2}{\sigma_k^2} + \sum_{\sigma_k > 0, k > M} \frac{|H_{\alpha/\sigma_k^p}^p(g_k) - g_k|^2}{\sigma_k^2}. \end{aligned}$$

For a given  $\epsilon > 0$  we choose  $M$  such that  $\sum_{\sigma_k > 0, k > M} |g_k|^2 / \sigma_k^2 < \epsilon$ . Since we can deduce from Lemma 5.1

$$|H_\alpha^p(x) - x| \leq |x|$$

we can estimate

$$\|R_\alpha^p(g) - K^+ g\|^2 = \sum_{\sigma_k > 0, k \leq M} \frac{|H_{\alpha/\sigma_k^p}^p(g_k) - g_k|^2}{\sigma_k^2} + \epsilon.$$

Furthermore, we see from Lemma 5.1

$$H_\alpha^p(x) \rightarrow x \quad \text{for } \alpha \rightarrow 0.$$

and hence, for sufficiently small  $\alpha$  we have

$$\|R_\alpha^p(g) - K^+ g\|^2 < 2\epsilon.$$

□

The above theorem does only proof convergence on the range of the operator. To obtain results on the speed of convergence one may assume special sparseness or decay properties similar to [3]. We are not going to pursue further in this direction since the case of the singular basis is of limited interest in practical applications. Moreover, convergence for noisy data has not been shown.

## 5.2 Regularization is impossible

In this section we present an example where a sparsity constraint with exponent  $p = 0$  does not lead to a regularization. In particular the minimization of the Tikhonov functional is not well-posed in the sense that it does not have a solution. To this end, we design an operator  $A$  which does not act well on a given orthonormal basis  $(\psi_k)$ . Let  $\{h_k\}$  be a countable set which is dense in the unit-ball of  $Y$ , i.e.  $\|h_k\|_Y = 1$  and for every  $g \in Y$  with  $\|g\|_Y = 1$  and every

$\epsilon > 0$  there is an index  $k_0$  such that  $\|g - h_{k_0}\|_Y \leq \epsilon$ . We define the operator  $A$  on the basis  $(\psi_k)$  by

$$A\psi_k = h_k, \text{ i.e. } Ku = \sum_k u_k h_k. \quad (20)$$

**Proposition 5.5.** *Let  $K$  be defined by (20),  $\|g\|_Y^2 > \alpha$  and let further  $g$  be not a multiple of  $h_k$  for every  $k$ . Then the functional*

$$\Psi(u) = \|Ku - g\|_Y^2 + \alpha \sum_k \text{sgn}(|u_k|)$$

*does not have a minimizer.*

*Proof.* Since the penalty term  $\sum_k \text{sgn}(|u_k|)$  does only depend on the number of coefficients we minimize separately over subspaces of a given dimension  $n$ .

As first case we consider  $n = 0$ , i.e. we minimize just over  $u = 0$ . We observe that  $\Psi(0) = \|g\|_Y^2$ .

As second case we observe that  $\Psi(u) \geq 2\alpha$  if  $u$  has more than two different non-zero entries.

The last case is to minimize over the one-dimensional subspaces  $X_k = \text{span}\{e_k\}$  where  $e_k$  is the canonical basis of  $\ell^2$ . The values of  $\Psi$  are

$$\Psi(d_k e_k) = \|d_k h_k - g\|_Y^2 + \alpha.$$

Since  $\{h^k\}$  is dense in the unit ball may take  $d_k = \|g\|_Y$  and find a sequence  $h_l$  such that  $\|g\|_Y h_l \rightarrow g$  for  $l \rightarrow \infty$ . Hence, the minimal value of  $\Psi$  over all subspaces  $X_k$  is  $\alpha$ , i.e.

$$\inf_{u \in \bigcup X_k} \Psi(u) = \alpha$$

and this infimum is not attained since  $g$  is not a multiple of a basis vector  $h_k$ .  $\square$

It is clear that a similar example can be constructed if the vectors  $h_k$  accumulate at a single point: take  $g$  as the accumulation point of  $h_k$ .

**Remark 5.6.** *We remark that also the constrained model*

$$\text{Minimize } \sum_k \text{sgn}(|u_k|) \text{ s.t. } \|Ku - g^\delta\|_Y \leq \epsilon \quad (P_\epsilon)$$

*is not well posed with  $K$  from (20) since it has an infinite number of solutions. One may say that this situation is a little better than that of Proposition 5.5 since now solutions are available. An easy example shows, that regularization need not to happen in this setup. Let  $g^+ = h_1$  and let  $\|g^+ - g^\delta\|_Y \leq \delta$ . The corresponding true solution is  $u^+ = e_1$ . Then there is a sequence  $h_l$  such that  $h_l \rightarrow h_1 = g^+$ . Moreover, for sufficiently large  $l$ ,  $u^{\epsilon, \delta} = e_l$  is a solution of  $(P_\epsilon)$  with  $\epsilon = \tau\delta$  with  $\tau > 1$  (assumed that the norm of  $g^\delta$  is not too small). Finally,  $\|u^{\epsilon, \delta} - u^+\| = \sqrt{2}$  is not converging to zero for  $\epsilon = \tau\delta$  and  $\delta \rightarrow 0$ .*

## 6 Conclusions

In this paper the regularizing properties of sparsity constraints have been analyzed. Special attention was paid to convergence rates in norm and to the source conditions. For  $1 < p \leq 2$  we could show, as a simple application of the results of Burger, Osher [7] and the inequality of Xu and Roach [23] (or the basic inequality in Lemma 2.3 from [4]), that a convergence rate  $\sqrt{\delta}$  in the 2-norm can be achieved by a source condition saying that  $u^+$  has to be in a weighted  $\ell^p$  space with small  $p$ , see Remark 3.5.

The case  $p = 1$  needed a special technique: the Bregman-Taylor-distance from [5]. Applying this, a convergence rate  $\sqrt{\delta}$  in the stronger 1-norm could be achieved under the source condition that  $u^+$  is finitely supported.

The incipient discussion on regularization with  $p < 1$  showed two things: First, regularization may or may not be possible and second, the regularization properties depend on the interplay of the operator  $A$  and on the choice of the basis functions  $(\psi_k)$ —a phenomenon which is not known for  $p \geq 1$ . One may conjecture that if the operator  $A$  acts well on the basis  $(\psi_k)$  (in the sense that the values  $\frac{\langle A\psi_k | A\psi_j \rangle}{\|A\psi_k\|_Y \|A\psi_j\|_Y}$  are not too large) regularization is possible. This would parallel observations in the framework of compressed sensing on the mutual coherence of dictionaries, see [11].

## References

- [1] Thomas Blumensath, Mehrdad Yaghoobi, and Mike Davies. Iterative hard thresholding and  $l^0$  regularisation. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2007.
- [2] Thomas Bonesky, Kristian Bredies, Dirk A. Lorenz, and Peter Maass. A generalized conditional gradient method for nonlinear operator equations with sparsity constraints. *Inverse Problems*, 23:2041–2058, 2007.
- [3] Lasse Borup and Morten Nielsen. Some remarks on shrinkage operators. Aalborg University, Dept. of Mathematical Sciences, 2003.
- [4] Kristian Bredies and Dirk A. Lorenz. Iterated hard shrinkage for minimization problems with sparsity constraints. *SIAM Journal on Scientific Computing*, 30(2):657–683, 2008.
- [5] Kristian Bredies and Dirk A. Lorenz. Linear convergence of iterated soft-thresholding. *To appear in Journal of Fourier Analysis and Applications*, 2008.
- [6] Kristian Bredies, Dirk A. Lorenz, and Peter Maass. A generalized conditional gradient method and its connection to an iterative shrinkage method. *To appear in Computational Optimization and Applications*, 2008.
- [7] Martin Burger and Stanley Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411–1420, 2004.

- [8] Matrin Burger, Elena Resmerita, and Lin He. Error estimation for Bregman iterations and inverse scale space methods in image restoration. *Computing*, 81(2–3):109–135, 2007. Special Issue on Industrial Geometry (Guest editors: B. Jüttler, H. Pottmann, O. Scherzer).
- [9] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications in Pure and Applied Mathematics*, 57(11):1413–1457, 2004.
- [10] Ingrid Daubechies, Massimo Fornasier, and Ignace Loris. Accelerated projected gradient method for linear inverse problems with sparsity constraints. To appear in *Journal of Fourier Analysis and Applications*, 2008.
- [11] David L. Donoho and Michael Elad. Optimally-sparse representation in general (non-orthogonal) dictionaries via  $\ell^1$  minimization. *Proceedings of the National Academy of Sciences*, 100:2197–2202, 2003.
- [12] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of Inverse Problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 2000.
- [13] Massimo Fornasier and Holger Rauhut. Recovery algorithms for vector valued data with joint sparsity constraints. *SIAM Journal on Numerical Analysis*, 46(2):577–613, 2008.
- [14] Roland Griesse and Dirk A. Lorenz. A semismooth Newton method for Tikhonov functionals with sparsity constraints. *Inverse Problems*, 24:035007 (19pp), 2008.
- [15] Bernd Hofmann, Barbara Kaltenbacher, Christiane Poeschl, and Otmar Scherzer. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems*, 23(3):987–1010, 2007.
- [16] Dirk A. Lorenz. Non-convex variational denoising of images: Interpolation between hard and soft wavelet shrinkage. *Current Development in Theory and Application of Wavelets*, 1(1):31–56, 2007.
- [17] Ronny Ramlau and Gerd Teschke. Tikhonov replacement functionals for iteratively solving nonlinear operator equations. *Inverse Problems*, 21(5):1571–1592, 2005.
- [18] Ronny Ramlau and Gerd Teschke. A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints. *Numerische Mathematik*, 104(2):177–203, 2006.
- [19] Elena Resmerita. Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Problems*, 21(4):1303–1314, 2005.
- [20] Elena Resmerita and Otmar Scherzer. Error estimates for non-quadratic regularization and the relation to enhancement. *Inverse Problems*, 22(3):801–814, 2006.

- [21] Frank Schöpfer, Alfred K. Louis, and Thomas Schuster. Nonlinear iterative methods for linear ill-posed problems in Banach spaces. *Inverse Problems*, 22:311–329, 2006.
- [22] Thomas Schuster, Peter Maass, Thomas Bonesky, Kamil S. Kazimierski, and Frank Schöpfer. Minimization of Tikhonov functionals in Banach spaces. *Abstract and Applied Analysis*, 2008:Article ID 192679, 19 pages, 2008.
- [23] Zong Ben Xu and Gary F. Roach. Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces. *Journal of Mathematical Analysis and Applications*, 157(1):189–210, 1991.